(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 January 2005 (20.01.2005)

PCT

(10) International Publication Number
WO 2005/006768 A1

(51) International Patent Classification[7]: H04N 9/00, 7/16, 7/00

(21) International Application Number: PCT/US2003/022562

(22) International Filing Date: 18 July 2003 (18.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/480,068 20 June 2003 (20.06.2003) US

(71) Applicant *(for all designated States except US)*: **NIELSEN MEDIA RESEARCH, INC** [US/US]; 770 Broadway, New York, NY 10003 (US).

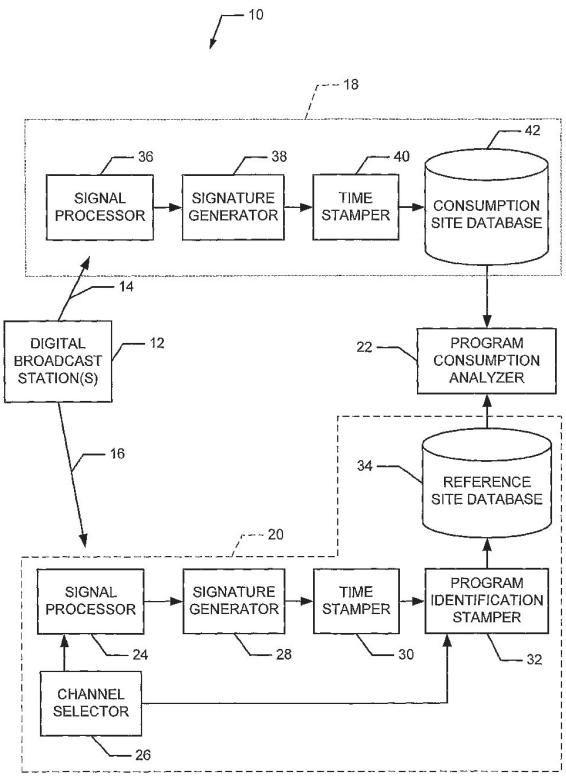
(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **LEE, Morris** [US/US]; 31790 U.S. 19 N. #187, Palm Harbor, FL 34684 (US). **SCHILP, William, H.** [US/US]; 455 Pennsylvania Avenue, Palm Harbor, FL 34683-5233 (US). **HICKS, Michael, A.** [US/US]; * (US). **SCHWERER, Ronald, G.** [US/US]; 517 Belle Isle Avenue, Belleair Beach, FL 33786 (US). **WRIGHT, David, H.** [US/US]; 3102 Bishop Drive, Safety Harbor, FL 34695 (US). **DENG, Kevin** [US/US]; 2106 Gull Lane, Safety Harbor, FL 34695 (US).

(74) Agent: **HANLEY, Mark, G.**; Grossman & Flight LLC, 20 N. Wacker Drive, Suite 4220, Chicago, IL 60606 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

[Continued on next page]

(54) Title: SIGNATURE-BASED PROGRAM IDENTIFICATION APPARATUS AND METHODS FOR USE WITH DIGITAL BROADCAST SYSTEMS



(57) Abstract: Signature-based program identification apparatus and methods for use with digital audio and/or video systems are disclosed. The disclosed systems and methods develop signatures of digital audio and/or video content from a digital data stream at two or more sites and compare the signatures to one another to identify the content.

WO 2005/006768 A1

AMR EX. 2027
Samsung v AMR
IPR2024-01348

(84) **Designated States** *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

SIGNATURE-BASED PROGRAM IDENTIFICATION APPARATUS AND METHODS FOR USE WITH DIGITAL BROADCAST SYSTEMS

FIELD OF THE DISCLOSURE

[0001] The present disclosure relates generally to digital broadcast systems and, more specifically, to signature-based audio/video content identification apparatus and methods for use with digital audio/video content delivery systems.

BACKGROUND

[0002] Digital audio/video content delivery and distribution systems have undergone rapid development and deployment in recent years due, at least in part, to their ability to efficiently distribute large numbers of high-fidelity video and/or audio content and programs as well as other digital services or content over a wide geographic region. For example, in the case of satellite-based digital television and wireless digital video/audio delivery systems, little, if any, land-based infrastructure (e.g., ground stations, cable lines, etc.) is needed to provide services to subscribers on a world-wide basis. As a result, such satellite-based and other wireless systems enable digital broadcasters and audio/video content providers to develop markets for their content, programs and/or services in remote and/or underdeveloped regions of the world in a cost-effective manner.

[0003] In general, regardless of whether a digital audio/video content delivery system uses one or more satellites, cable transmission lines, ground-based wireless transmission stations, the Internet or any combination thereof, such a system enables users to receive high-fidelity video and/or audio content, as well as user-interactive features or tools such as, for example, pictographic program guides, grid-based guides, etc. that enable users to plan and perform program consumption activities.

More specifically, digital audio/video content delivery systems that deliver digital television may transmit one or more high-bandwidth signals, each of which is typically composed of a stream of data or data packets having a plurality of video, audio and/or other digital programs multiplexed therein. Users or subscribers typically have a digital decoder unit which, in the case of a digital video or television system, is commonly referred to as a "set-top box." The set-top box can, based on subscriber commands, selectively de-multiplex data packets associated with a particular program from a particular data stream, decode those data packets and send a signal to a television or video monitor for viewing. Some digital televisions provide integrated or built-in set-top box functionality and, thus, do not require a separate set-top box.

[0004] Of course, as is the case with analog content providers and broadcasters, digital content providers and broadcasters (e.g., digital television, audio, etc.) desire to meter (i.e., monitor) the consumption of the programs and/or other services provided to users or subscribers to thereby generate audience program consumption behavior or patterns. Such audience consumption behavior or patterns may be used to develop program ratings information, to improve targeted marketing or advertising campaigns, to improve the types of programs and/or services and the times at which those programs and/or services are delivered, to assess the value of a particular time slot and program for purposes of determining what advertisers should pay for usage of that time slot., etc.

[0005] A metering system designed for use with an analog content delivery system, such as an analog television system, generates viewing records or other viewing information that is typically collected from a group of statistically selected households. Each of the statistically selected households has a data logging and

processing unit commonly referred to as a "home unit" or "site unit." The site unit is usually coupled to a variety of attachments that provide inputs to the site unit or that receive outputs from the site unit. For example, a frequency detector may be used to detect the local oscillator frequency of the television to identify the channel to which the television is currently tuned. Also, for example, a people counter, may be located in the viewing space of the television and in communication with the site unit, thereby enabling the site unit to actively or passively detect the identities of the persons currently viewing programs displayed on the television. In some cases, the metering system is designed to capture signature information associated with audio and/or video content that has been consumed in the household.

[0006] The site unit usually processes its inputs, including the inputs from its attachments, to produce viewing records. In general, the viewing records typically contain program identification information (i.e., information that can be used to identify a particular program, such as channel information obtained from the frequency detector and signature information) and audience information such as the numbers and/or identities of people viewing that particular program obtained from the people counter. The viewing records may be generated on a periodic basis (i.e., at fixed time intervals) or may be generated in response to a change in an input such as, for example, a change in the identities or number of persons viewing the television, a change in the channel tuning information (i.e., a channel change), etc. Of course, viewing records may contain any other information such as the identities of viewers present at the viewing time.

[0007] Often, the site unit collects a quantity of viewing records and transmits collected viewing records, usually daily, to a central office or data processing facility for further processing or analysis. The data processing facility receives viewing

records from site units located in some or all of the statistically selected households and additionally receives information from one or more reference sites. The reference sites are typically configured to capture information associated with the audio/video content broadcast to specific content distribution areas. For example, a reference site may collect signatures for all of the audio and/or video content distributed within its assigned area and then transmit the collected signatures to the data processing facility. The data processing facility analyzes the viewing records to ascertain the viewing behaviors of a particular household or group of households selected from all participating households. Additionally, the central data processing facility may generate viewing behavior statistics and other parameters indicative of viewing behavior associated with all of the participating households.

[0008] To generate viewing behavior information from viewing records, the central office or data processing facility may compare each of the signatures generated by a household unit to the signatures generated by the reference site. When a signature from the household unit matches a signature from a reference site, the identity of the content associated with matching signatures may be determined using data such as a list of content (i.e., a schedule lineup of television programming or television guide) associated with the reference site signatures and/or the time at which, and the channel from which, the signature was captured at the reference site. Such a matching process may be carried out for each of the viewing records received by the data processing facility, thereby enabling the data processing facility to reconstruct what programs were watched by all participating households and the times at which the programs were watched. Alternatively, such a matching process may be used only when other data, such as channel data or time data, is unavailable.

[0009] Signature information may be used in the foregoing manner to perform content identification because each signature is designed to uniquely correspond to the content from which it was collected. Thus, a match between two signatures indicates that the matching signatures were derived from the same content. While a variety of signature generation and capture techniques are well known in the art, such techniques are suited for use with audio/video content in an analog format.

BRIEF DESCRIPTION OF THE DRAWINGS

[0010] Fig. 1 is a block diagram of an example system that uses signatures to identify digital audio/video content.

[0011] Fig. 2 is flow diagram of an example manner in which the system shown in Fig. 1 may be used to identify digital audio/video content consumed at a consumption site.

[0012] Fig. 3 is a diagrammatic view of an example data stream that may be used with the example system shown in Fig. 1.

[0013] Fig. 4 is a diagrammatic view of an example manner in which content reference signature information may be generated using an example digital data stream.

[0014] Fig. 5 depicts an example data structure that may be used to store reference signature information generated using the technique shown in Fig. 4 for a plurality of programs.

[0015] Fig. 6 is a diagrammatic view of an example manner in which sets of data stream signatures may be generated at the consumption site shown in Fig. 1.

[0016] Fig. 7 is an example data structure that may be used to store consumption site signature information generated using the technique depicted in Fig. 6.

[0017] Fig. 8 is an example digital audio/video content delivery metering system that may use the signature-based program identification apparatus and methods described herein.

DETAILED DESCRIPTION

[0018] Fig. 1 is a block diagram of an example system 10 that uses signatures to identify digital audio and/or video content. For purposes of illustration only, the system 10 is portrayed as a digital television broadcast system. However, the system may be implemented as any type of system configured to deliver audio/video content remotely and/or locally for consumption at a consumption site. The system 10 may include one or more digital broadcast stations 12 that transmit video programs, audio programs and/or other digital content or programs (e.g., television programs, advertisements, etc.) via communication paths or links 14 and 16 to at least one consumption site 18 and at least one reference site 20. The digital broadcast stations 12 may include one or more television stations, radio stations, Internet providers or any other transmission sites capable of transmitting video and/or audio content in a digital format. The communication paths or links 14, 16 may be implemented using any desired combination of wireless and hardwired links, including, for example, satellite links, cellular links, telephone lines, cable, Internet, a digital subscriber line, WiFi, Ethernet, etc. Of course, in the case where the digital broadcast stations 12 are transmitting television programs, the digital signals conveyed via the links 14, 16 contain both video and audio information and, in some cases, may also contain other information for providing a graphical program guide and/or other user-interactive features. In the case where the digital broadcast stations 12 are transmitting a plurality of programs simultaneously, the program information may be transmitted using a plurality of separately modulated signals (i.e., independent signals), using one

or more multi-program signals (i.e., signals within which multiple programs are multiplexed), or any combination thereof. For illustrative purposes only, the system 10 described herein and the examples described with respect thereto refer to broadcast content and/or broadcast programs. Broadcast content, as used herein, refers to content that is widely disseminated via any transmission method including any of the methods described above. However, the system described herein is not limited to use with systems that deliver and receive broadcast content. To the contrary, the system described herein is more generally applicable and, thus, is also suitable for use with systems that deliver and receive local content. For example, a television may be used to consume digital content supplied via a local source such as a digital video device (DVD). The system 10 described herein can be used to generate signatures and identify content delivered locally as well as content delivered via a broadcast system and can be used to generate signatures and identify content that was delivered via a broadcast system and then recorded for playback at a later time via a local device, such as a video cassette player, a digital video recorder, etc. Of course, signatures generated using content delivered via a local source can only be identified if the data processing facility has reference signatures that correspond to such locally delivered content. Signatures that correspond to locally delivered content can typically be obtained by using publicly released content that is usually consumed via a local device such as movies available on video cassette or digital versatile disks, musical recordings available on compact discs, movies and/or musical recordings downloadable via the Internet, etc.

[0019] The consumption site 18 shown in Fig. 1 is typically associated with the viewing location of a person subscribing to or otherwise enabled to access the services provided by one or more of the digital broadcast stations/services 12 although

it may also be the site of a digital television having access to a non-subscription broadcasting system/service. The consumption site 18 may be the home of an individual, a business location, including, without limit, any location, indoor or outdoor, at which content may be consumed. Although one consumption site (i.e., the consumption site 18) is shown in Fig. 1 for purposes of simplifying the following description, a plurality of other such consumption sites typically have access to the signals transmitted by the digital broadcast stations 12. Likewise, while a single reference site 20 is shown in Fig. 1, other such reference sites could be included within the system 10.

[0020] While the reference site 20 is described below as having a single signal processing channel (i.e., the ability to process content delivered via a single channel at any given time), multiple signal processing channels are typically employed. Thus, the reference site 20 is typically configured to process information associated with multiple channels and multiple content simultaneously, thereby enabling the reference site 20 to generate a library of content identification information. In light of the exemplary system 10 being described herein as a digital television broadcast system, the content delivered thereby is hereinafter referred to as television programs or programs. However, as described above, the system 10 is not limited to processing television programs to generate signatures but may instead be used to process any type of digital audio content and/or digital video content. Each of the signal processing channels may be tuned to a particular channel or program so that information associated with each of those tuned channels or programs may be processed simultaneously to develop a library of reference information (e.g., signature information) over a given time period (e.g., a rolling twenty-four hour period).

[0021] In general, the example reference site 20 shown in Fig. 1 receives signals from the digital broadcast stations 12 and selectively processes one or more broadcast program data streams contained within those signals to generate a series of reference signatures for each program data stream that is selected for processing. The reference signatures are generated by using or extracting raw binary information contained within the digital program data streams so that, within a particular time interval, each series of reference signatures is substantially uniquely characteristic of the program data stream from which the reference signature generation information was extracted. Thus, as will be described in greater detail below, the reference signature information may be used to later identify broadcast programs from which the reference signature information was derived. For example, a reference signature generated using the information contained within a known television program signal may be used to identify that television program within an unknown or unidentified signal. More specifically, if the unknown or unidentified signal contains a signature which substantially matches the reference signature information associated with a known program, it can be reliably determined that the unknown or unidentified signal contains the known program.

[0022] The consumption site 18 is located within the same broadcast region as the reference site 20 and, therefore, receives the same broadcast signals that are received at the reference site 20. The consumption site 18 generates a series of signatures in a manner similar or identical to that of the reference site 20 for the program currently being consumed (e.g., viewed, listened to, etc.) at the consumption site 18. However, as the subscriber or user at the consumption site 18 changes the currently selected program, the series of signatures generated by the consumption site 18 changes in a corresponding manner. Thus, the series of signatures generated by the consumption

site 18 may include a plurality of distinct signature sequences, each of which may correspond to a different program, and each of which may have a length corresponding to the amount of time spent consuming that program (e.g., the amount of time during which the consumption site 18 was tuned to that program).

[0023] As described in greater detail below, a program consumption analyzer 22 receives the series of signatures generated by the consumption site 18 (and other consumption sites in the event more than one consumption site is used within the system 10) and compares the signatures therein to the reference signatures generated by the reference site 20. Typically, the program consumption analyzer 22 does not receive any program identification information along with the series of signatures received from the consumption site 18 because the consumption site 18 is not typically configured to generate such information. The program consumption analyzer 22 can compare the series of signatures generated by one or more consumption sites such as, for example, the consumption site 18, to reference signatures associated with known programs to ascertain the identities of programs viewed and the times at which the programs were viewed at the consumption site 18.

[0024] Now turning in detail to the reference site 20 shown in Fig. 1, a signal processor 24 extracts data associated with a particular channel as determined by a channel selector 26. The signal processor 24 can use known techniques to demodulate, demultiplex (if a multi-program signal has been demodulated), and decode program data received via the link or communication path 16. Some or all of the data extracted by the signal processor 24 may be audio data sent in conjunction with video information such as, for example, digital television signals, advertisements, etc. Alternatively, some or all of the data extracted and processed by the signal processor 24 may be video data or information. In the case where the

signal processor 24 receives digital television signals from the digital broadcast stations 12, the channel selector 26 controls the signal processor 24 to extract data packets having a particular identifier (e.g., in the data packet header) associated with a particular broadcast channel.

[0025] Each of the broadcast stations 12 may transmit one or more digital data streams, each of which may be associated with a particular major channel. In turn, each major channel may include two or more minor channels or sub-channels, each of which may include a video data stream and one or more audio data streams. The reference site 20 retains program lineup information that supplies tuning information to the channel selector 26. For example, the lineup information may provide packet identifiers for each of the major and minor channels, video data streams, audio data streams, etc. along with program identification information (e.g., program titles or names) for each of the channels indexed over time. In this manner, the channel selector 26 can be configured using the program lineup information to cause the signal processor 24 to select video and/or audio data packets having identifiers associated with a particular major and/or minor channel. Typically, in the case where more than one signal processing channel is employed within the reference site 20, each processing channel has a channel selector configured to cause its corresponding signal processor to select data packets associated with a different channel. In that case, as noted above, the reference site 20 can be configured to generate signature information simultaneously for a plurality of channels for a predetermined period of time (e.g., a rolling twenty-four hour period).

[0026] In some example systems, one or more broadcast programs may be associated with more than one audio data stream. For example, multiple audio data streams, each of which is associated with a different language, may be associated with

a single video program. In cases where a broadcast channel or program includes a plurality of alternative audio data streams, the signal processor 24 may selectively extract or parse data packets associated with one, some or all of the available audio data streams.

[0027] The signal processor 24 sends extracted or parsed data packets associated with the currently selected channel to a signature generator 28. The signature generator 28 uses information within the extracted data to generate a series of signatures substantially uniquely characteristic of the selected program or channel data stream(s). In examples where the currently selected channel includes a plurality of alternative audio data streams, the signature generator 28 may receive data packets associated with one, some or all of those audio data streams from the signal processor 24 and generates a series of signatures for each of the data streams. As described in greater detail in connection with Figs. 2 and 3, the signature generator 28 may combine or otherwise use portions of data packets to generate signatures. In particular, the signature generator 28 may concatenate cyclical redundancy check (CRC) values associated with the data packets in a manner that generates a series of digital values (i.e., signatures) that is substantially unique to the program data stream currently being extracted by the signal processor 24.

[0028] The reference site 20 also includes a time stamper 30 and a program identification stamper 32. The time stamper 30 associates a time value with each signature value generated by the signature generator 28. The program identification stamper 32, based on the channel currently selected by the channel selector 26, uses program lineup information and the time stamp values to identify the program currently being processed by the signal processor 24. In addition, the program identification stamper 32 associates the information pertaining to the identity of the

program currently being processed with the signature information generated by the signature generator 28. Each generated reference signature is stored along with its time stamp and program identification information within a reference site database 34.

[0029] Reference signature information may be generated for multiple programs and, in some cases, multiple audio/video data streams for one or more of the multiple programs, and stored within the reference site database 34. For example, the signal processor 24 and channel selector 26 could be configured to select different programs or channels at different times for processing. In that case, the reference site database 34 contains reference signatures for different programs, where the reference signature information generated for each program was generated at a different time.

[0030] Additionally or alternatively, as noted above, the reference site 20 may be configured to provide multiple processing channels so that reference signature information can be generated for a plurality of programs simultaneously. In that case, the reference site 20 may include an additional signal processor 24, channel selector 26, signal generator 28, time stamper 30 and/or program identification stamper 32 for each additional processing channel. In the case where the reference site 20 includes multiple processing channels, each of the channel selectors 26 is typically configured to cause its corresponding signal processor 24 to extract data packets associated with a different major and/or minor broadcast channel. Also, once configured, the channel selectors 26 are typically not reconfigured and, thus, control their respective channels to extract data packets associated with a particular channel on an ongoing basis.

[0031] The consumption site 18 includes a signal processor 36, a signature generator 38, a time stamper 40 and a consumption site database 42, all of which may be similar to the signal processor 24, the signature generator 28, the time stamper 30 and the reference site database 34, respectively. However, as depicted in the example

system 10 of Fig. 1, the consumption site 18 does not have blocks similar to the program identification stamper 32 and the program selector 26. While the consumption site 18 may employ an apparatus that enables a user or subscriber to control program selection, any such program selection apparatus does not output information that can be readily used to extract and store currently selected program identification information in the consumption site database 42.

[0032] The time stampers 30 and 40 are preferably substantially synchronized so that if the signature generators 28 and 38 generate signatures at substantially the same instant of time, the time values generated by the time stampers 30 and 40 and associated with those signatures are substantially the same. However, as described in greater detail below, absolute synchronization of the time stampers 30 and 40 is not required if the program consumption analyzer 22 is configured to look for signature matches in the reference site database 34 within a window (i.e., a range) of time surrounding a suspected time at which a consumption site signature was generated.

[0033] Any or all of the signal processors 24, 36, the signature generators 28, 38, the time stampers 30, 40, the program identification stamper 32, the channel selector 26, the reference site database 34, the program consumption analyzer 22 and the consumption site database 42 may be implemented using any desired combination of hardware, firmware and/or software. For example, each of the reference site 20, the consumption site 18 and the program consumption analyzer 22 may utilize a computer system or other processor system that executes machine readable instructions or programs to implement the apparatus and methods described herein. The machine readable instructions or programs may be embodied in software stored on a tangible medium such as a CD-ROM, a floppy disk, a hard drive, a digital versatile disk (DVD), and/or a memory.

[0034] The databases 34, 42 may be implemented using any known mass storage device(s) such as, for example, magnetic and optical disk drives. In addition, the databases 34, 42 may be combined into a single or common database, which may be embodied on a single or multiple mass storage devices, accessible by the program consumption analyzer 22.

[0035] Fig. 2 is flow diagram of an example manner in which the system 10 shown in Fig. 1 may be used to identify digital broadcast programs consumed at the consumption site 18. The signal processors 24, 36 acquire data associated with a program selected for consumption by a user at the consumption site 18 (blocks 70, 71). For example, the signal processors 24, 36 may extract or parse audio and/or video data associated with a particular program from a multi-program digital television signal sent by the digital broadcast stations 12 (blocks 70, 71). The signature generators 28, 38 then generate signature information by, for example, concatenating CRCs or portions of CRCs associated with the extracted data (blocks 72 and 73). Each signature may, for example, be generated using a portion of one or more CRCs in a manner that results in a substantially unique digital value that is substantially uniquely characteristic of the extracted data and, thus the program, currently being consumed at a particular time. In general, generating a signature composed of a large number of bits increases the probability that the signature is unique to the program currently being consumed (i.e., decreases the probability that another program, if consumed at the same time, would result in the same signature). However, signatures having a large number of bits increases processing overhead and, thus, may decrease the overall processing speed and efficiency of the system 10 (Fig. 1). Each signature generated by the signature generators 28, 38 is stored along with a time stamp generated by the time stampers 30, 40 (block 74).

[0036] The program consumption analyzer 22 (Fig. 1) extracts the signature information or data (along with associated time stamps) stored within the consumption site database 42 and compares this signature information to the reference signature information stored within the reference site database 34 during a relevant time period (block 76). In particular, the program consumption analyzer 22 examines the time stamp values associated with the signatures from the consumption site database 42 that are currently being analyzed to reference signature information stored within the reference site database 34 during a relevant range of times or window of time. Due to unequal signal transmission delays between the digital broadcast stations 12 and the sites 18 and 20, the program consumption analyzer 22 may compare the signature information from the consumption site database 42 that is currently being analyzed to reference site information that was acquired over a somewhat longer amount of time (i.e., a wider or larger time window). For example, the comparison may extend several seconds prior to and following the time associated with the earliest and latest time stamps of the consumption site signature information being analyzed, thereby eliminating the need for absolute or precise synchronization between the sites 18 and 20 (e.g., the time stampers 30 and 40).

[0037] Because the user at the consumption site 18 can change the program being consumed (e.g., by changing channels), the program consumption analyzer 22 analyzes the series of signatures provided by the consumption site database 42 using sets or blocks of signatures representative of consumption activity over a relatively brief time interval such as, for example, thirty seconds. Of course, any other time interval could be used instead.

[0038] For each comparison made, the program consumption analyzer 22 determines whether a substantial match of the consumption site signature information

has been found within the reference site database 34 (block 78). If a substantial match has been found, the program consumption analyzer 22 uses the program identification information stored along with the reference signature information to identify the program that was consumed within the consumption site 18 (block 80). In the event that a match is not found, any desired non-match handling routines may be used to determine and initiate an appropriate action (block 82). For example, such non-match handling routines may include a window adjustment routine to enable adjustment of the relevant time window used by the program consumption analyzer, may display or otherwise output a closest match for final determination by, for example, a reference site operator.

[0039] Fig. 3 is a diagrammatic view of an example data stream 100 that may be used with the example system 10 shown in Fig. 1. The example data stream 100 includes a plurality of data packets 102, 104 and 106 having respective payload or data portions 108, 110 and 112, respective CRC portions 114, 115 and 116 and respective header portions 117, 118 and 119. The payload portions 108, 110 and 112 include information related to audio and/or video content that may be presented to a user via an output device. On the other hand, the CRC portions 114, 115 and 116 and header portions 117, 118 and 119 are non-payload digital information.

[0040] The data stream 100 may be compliant with any known digital transmission protocol or standard. For example, in the case where the digital broadcast stations 12 are transmitting digital television signals that are compliant with the Advanced Television Systems Committee (ATSC) standard, the data stream 100 may be a Dolby AC-3 compliant digital audio signal. By way of background, the Dolby AC-3 standard is a compressed multi-channel digital audio standard that was developed by the Dolby Corporation and that was adopted by the ATSC for use with digital

television transmissions within the United States of America. In any event, if the data stream 100 is an AC-3 compliant data stream, the data packets 102-106 are transmitted at a rate of 31.25 per second (i.e., a packet every 32 milliseconds) and the CRCs 114-116 within the data packets 102-106 are sixteen-bit values.

[0041] The digital data stream 100 may instead be compliant with other protocols or standards such as, for example, the European Digital Video Broadcast (DVB) standard, the Japanese Terrestrial Integrated Services Digital Broadcasting (ISDB-T) standard, Motion Picture Expert Group – 2 (MPEG-2) standard, MPEG-1, MPEG-3, MPEG-4 Advanced Audio Code or any other version of MPEG, or any other similar or different digital video and/or audio transmission standard or protocol. In general, any digital data stream that provides audio and/or video information that can be extracted from the data stream can be used to implement the signature-based program identification apparatus and methods described herein. Thus, any digital data stream that can be used to generate unique signatures may be used. By way of example, any digital data stream having a substantially random sequence of digital values, e.g., CRCs or any other predetermined data packet portions that are computationally related to data packet content, may be used to generate unique signatures. More specifically, predetermined data packet portions that are computed based on other portions of data packet content (e.g., CRCs) can provide sufficiently random sequences of digital values that may be used to implement the apparatus and methods described herein. It should be understood that while the sequences of digital values generated by the apparatus and methods described herein are, numerically speaking, substantially random, they are computationally deterministic with respect to the channel and program data packets from which they are derived. As a result, the sequences of digital values generated using the apparatus and methods described

herein are substantially uniquely characteristic of the content from which they are derived.

[0042] Further, it should be understood that both broadcast station characteristics and program content can contribute to the uniqueness of the digital sequences generated using the apparatus and methods described herein. Thus, in the event that a program is transmitted simultaneously (e.g., simulcast) by two or more broadcast stations, the digital sequences generated by two or more receiving stations (e.g., reference sites and consumption sites) for two or more simulcast versions of a program using the apparatus and methods described herein can, nevertheless, be different due to differences in the characteristics of the broadcasting stations. For example, it is highly improbable that two broadcast stations encoding the same program will encode (e.g., compress) the program content in precisely the same manner to generate identical digital data packets. Thus, if CRCs, or any other predetermined data packet portions that are computationally related to other data packet content, are used to form signatures for two or more versions of a simulcast program, each version of the simulcast program will yield a different, distinguishable sequence of digital values.

[0043] Fig. 4 is a diagrammatic view of an example manner in which program reference signature information may be generated using an example digital data stream 120. The example digital data stream 120 includes a plurality of data packets 122, 124, 126, 128 and 130 having respective data portions 132, 134, 136, 138 and 140, respective CRCs 141, 142, 143, 144 and 145, and respective header portions 146, 147, 148, 149 and 150. The example data stream 120 is associated with a particular type of program (e.g., a digital television program).

[0044] As depicted in the example of Fig. 4, the signature generator 28 (Fig. 1) forms signatures for selected ones of the data packets 122-130 (as opposed to all of the data packets) to reduce the amount of reference signature information generated and, thus, the amount of reference signature information to be stored in the reference site database 34. Although the example of Fig. 4 depicts signatures being generated for half the data packets (i.e., by skipping every other data packet), more or fewer data packets could be skipped (i.e., a signature may be generated for every X data packets) if desired or, alternatively, signatures could be generated for every data packet. In general, as the number of skipped data packets increases, the amount of signature information generated for the program associated with the data stream 120 decreases. However, as will been seen in connection with the following discussion, as the amount of reference signature information generated decreases, the amount of signature information generated by the consumption site 18 may have to be significantly increased to guarantee that the signature information generated at the consumption site 18 will include at least one or two reference signature matches.

[0045] The signature generator 28 generates a signature S(N) by combining portions of the CRCs 143 and 145. Such a combination may involve a concatenation of the entire CRCs 143 and 145, a concatenation of equal size or unequal size portions of the CRCs 143 and 145, a mathematical or logical operation (e.g., a summation, division, multiplication, ANDing, ORing, etc.) using the CRCs 143 and 145 (or portions thereof) as operands, or some other combining technique. One reference signature generation technique that is particularly useful in the case where the example data stream 120 is an AC-3 compliant audio data stream concatenates the CRCs of data packets that are a predetermined number of data packets apart. With this reference signature generation technique, the signature S(N) of a data packet N is

mathematically given by Equation 1 below where the value "X" is the number of data packets skipped between signature generations.

[0046] Equation 1 $S(N) = CRC(N+X) * 2^{16} + CRC(N)$

[0047] Thus, if the number of data packets skipped between signatures equals thirty-two, the generation of signatures starting with a data packet "1" according to Equation 1 above results in the generation of signatures S(1), S(33), S(65), S(96).... More generally, the use of Equation 1 as described above generates signatures S(N) using a sliding window of two CRCs that are X data packets apart. This sliding window is shown in two successive positions represented by the brackets labeled S(N) and S(N+X), respectively, of Fig. 4.

[0048] Of course, the reference signature generation technique represented diagrammatically in Fig. 4 is only one example and many other variations of the technique shown thereby may be employed with similar results. For example, the number of data packets skipped (i.e., value X) may be made greater than or less than thirty-two. Further, the number of bits in the resulting signatures may be less than or greater than thirty-two. For example, four bits of the data packet N+X may be concatenated with the full sixteen bits of the data packet N to form a twenty bit signature. Still further, bits from more than two CRCs may be used to form each signature, in which case, signatures could have more than thirty-two bits. In the case where three or more CRCs are combined, the sliding window can be envisioned as having three or more corresponding prongs. Still further, in the case where three or more CRCs are used to generate signatures, the CRCs may be evenly spaced within the data stream so that there are equal numbers of skipped data packets between each of the CRCs. Alternatively, the CRCs may be spaced in an unequal manner so that

the number of skipped data packets between one pair of CRCs used to generate a signature is greater than or less than the number of skipped data packets between another pair of CRCs used to generate that same signature.

[0049] Fig. 5 depicts an example data structure 170 that may be used to store reference signature information generated using the technique shown in Fig. 4 for a plurality of programs 172, 174 and 176. The data structure 170 shown in Fig. 5 depicts signature information for three programs (i.e., programs A, B and C) being generated substantially simultaneously. In other words, population of the data structure 170 with reference signature information requires the reference site 20 (Fig. 1) to have multiple signal processing channels that simultaneously capture or extract data packets associated with three programs. The reference signature values are stored in the data structure 170 according to the program with which they are associated (i.e., Program ID) and the times at which the data packets to which the reference signature value corresponds were acquired. In examples where reference signature information is generated using the technique described in connection with Fig. 4, the data stream 120 is an AC-3 compliant audio data stream, and the sliding window skips thirty-two packets (i.e., X=32), the times T(N), T(N+X), etc. are one second apart, and the reference signatures are generated using pairs of CRCs that are thirty-two packets (i.e., one second) apart (i.e., $S_A(N)$, $S_A(N+32)$, $S_A(N+64)$, etc.)

[0050] When comparing consumption site signature information (retrieved from the consumption site database 42 shown in Fig. 1) to reference site signature information, which may be stored in the reference site database 34 (Fig. 1) using a data structure similar or identical to the data structure 170 (Fig. 5), the program consumption analyzer 22 uses the time data (i.e., data associated with the columns T(N), T(N+X), etc.) to limit the comparison to a relevant time window. For example,

if the signature information processed by the program consumption analyzer 22 includes signatures generated at times T(N+X) and T(N+2X), the program consumption analyzer 22 may compare all of the reference signatures associated with the columns T(N+X) and T(N+2X) with the consumption site signatures currently being processed.

[0051] Although the time stampers 30, 40 are preferably substantially synchronized to maintain approximately the same absolute time, transmission delay times may result in the reference site 20 and the consumption site 18 processing data packets and generating signatures therefrom at somewhat different times. As a result, the time stamps accompanying consumption site signatures may not precisely match the time stamps associated with corresponding reference signatures (e.g., T(N), T(N+X), etc. of the data structure 170 of Fig. 5) and, in some cases, could be time shifted several seconds or more. Thus, when performing comparisons, the program consumption analyzer 22 may use a tolerance or delay error window to determine if a time stamp associated with the consumption site 18 corresponds to a time stamp associated with the reference site 20, but for a difference in data extraction time and/or transmission delay time. For example, if a time stamp associated with a consumption site signature is an absolute time T, the program consumption analyzer 22 may look for signature matches within the reference site database 34 (within, for example, a data structure similar or identical to the database 170 of Fig. 5) over a time window of $T-W_1$ to $T+W_2$, where W_1 may be, for example, three seconds and W_2 may be, for example, five seconds. Of course, other values for W_1 and W_2, including equal values, could be used instead.

[0052] Before discussing additional example techniques for generating consumption site signatures, it should be recognized that to enable exact matching of

signatures between the reference site 20 and the consumption site 18, individual signatures are generated in a consistent manner at the reference site 20 and the consumption site 18. Thus, for example, if the reference site 20 generates signatures by concatenating pairs of CRCs that are thirty-two data packets apart, then the individual signatures generated at the consumption site 18 are also generated by concatenating pairs of CRCs that are thirty-two data packets apart.

[0053] Although the techniques used at the consumption site 18 and the reference site 20 for generating individual signatures are consistent or the same, the sets of signatures generated by the consumption site 18 and the reference site 20 do not necessarily have to be the same. In other words, the consumption site 18 may generate sets of signatures that include signatures associated with data packets for which the reference site 20 has not generated signatures. Likewise, the reference site 20 may generate signatures for data packets for which the consumption site 18 has not generated signatures. Of course, the sets of data packets for which signatures are generated by the consumption and reference sites 18 and 20 have to overlap sufficiently (i.e., a sufficient number of matching signatures) to enable the program consumption analyzer 22 to reliably identify programs consumed at the consumption site 18 based on the reference signature data.

[0054] Fig. 6 is a diagrammatic view of an example manner in which sets of signatures associated with a program data stream may be generated by the consumption site 18 shown in Fig. 1. In general, the technique depicted in Fig. 6 repeatedly generates signatures for blocks or sets of consecutive data packets 200 and 202 separated by a delay 204. Each of the signatures within the sets 200 and 202 is generated in the same manner as individual signatures generated at the reference site 20. For example, in the case where the reference signatures are generated using the

technique depicted in Fig. 4, each of the signatures within the sets of signatures 200 and 202 is generated using pairs of CRCs that are thirty-two packets apart.

[0055] In addition, each of the sets of signatures 200 and 202 is of a size and, thus, extends over a time period, that guarantees that at least one reference signature corresponding to a data packet for which a signature was generated by the consumption site 18 and included in the set would have been generated by the reference site 20 during the same time period. As indicated above, the sets of signatures 200 and 202 are composed of signatures for consecutive data packets (i.e., no data packets are skipped within a set). Thus, in a case where the reference site 20 generates signatures using the technique depicted in Fig. 4 using pairs of CRCs that are thirty-two packets apart and skipping thirty-two packets between signatures, the sets 200 and 202 include at least thirty-two consecutive signatures, which requires the acquisition of sixty-four consecutive data packet CRCs. In this manner, each of the sets 200 and 202 could include at least one signature corresponding to a reference signature.

[0056] The duration of the delay 204 may be selected to correspond to the nature of program consumption behavior (e.g., channel changing behavior of users). For example, the delay 204 may be selected to be about thirty seconds or any other desired amount of time that enables the signature information gathered at the consumption site 18 to accurately depict the program consumption activities at the consumption site 18. Further, the delay 204 may also be selected to minimize, in view of the accuracy needed in connection with identifying program consumption activities, the amount of data extraction (e.g., CRC extraction), signature generation, signature data storage and comparison activity performed. In general, as the duration of the delay 204 increases, the accuracy (i.e., the resolution) with which program

consumption activities can be determined decreases and the amount of data processing and storage requirements decreases.

[0057] Fig. 7 is an example data structure 220 that may be used to store consumption site signature information using the technique depicted in Fig. 6. In general, the example data structure 220 is formed as an array of signature values (i.e., S(N), S(N+1), etc.) and associated time stamps (i.e., T(N), T(N+1), etc.). First and second signatures sets 222 and 224 correspond to the signature sets 200 and 202, respectively, of Fig. 6. By way of example, each of the signature sets 222 and 224 includes thirty-two signature values. However, due to the delay 204 (Fig. 6), signature values for 864 data packets are skipped in this example and, as a result, the signatures for data packets N+32 to N+927 are not generated. In the case where the data stream provides data packets at a rate of 31.25 per second (e.g., an AC-3 audio data stream), 864 skipped packets corresponds to about twenty-seven seconds.

[0058] In addition to the example signature generation methods discussed in connection with Figs. 4-7, other methods of generating signature information within the consumption and references sites 18 and 20 are possible. For example, in another example method, signatures may be formed for every m^{th} CRC within data streams processed by the consumption site 18 and every n^{th} CRC within data streams processed by the reference site 20, where m and n are relatively prime. In this manner, matching signatures will only be generated at every m*n packets. Each of the signatures may be, for example, composed of twenty bits taken from two consecutive CRCs. However, any other number of consecutive CRCs and/or bits may be used instead.

[0059] In still another example method, reference signature information may be generated at the reference site 20 by dividing CRCs by a predetermined whole

number and storing only those CRCs that are evenly divisible by the whole number along with the number of CRCs since the last evenly divisible CRC was stored in the reference site database 34. Thus, under this method, the reference site database 34 includes data pairs including the number of CRCs since the last CRC evenly divisible by the predetermined whole number was recorded and an associated CRC value. With this example method, data pairs are generated in the same manner as in the reference site 20 over a period of time sufficient to enable reliable comparisons to be performed. Under this method it is possible for a relatively large number of CRCs to be skipped (i.e., that are not evenly divisible by the predetermined whole number). As a result, it is possible to select a maximum number of CRCs that are allowed to elapse. When this maximum number of CRCs has elapsed, regardless of whether the current CRC is divisible by the predetermined whole number, the current CRC value is stored along with the number of CRCs occurring since the last CRC was stored.

[0060] More generally, signatures may be formed using one or more predetermined portions of the data packets composing a digital program stream. For example, rather than CRCs (or portions thereof), signatures may be generated using portions of data packets that are offset a predetermined number of bits from the header portion of the data packets. Additionally, signatures may be formed using all or part of the digital metadata that may form a portion of the digital audio or digital video content.

[0061] Fig. 8 is an example digital broadcast system 300 that may use the signature-based program identification apparatus and methods described herein. The digital television system 300 includes a digital broadcast station 302 that receives digital video and/or audio programs from a plurality of digital content providers 304 and 306. The digital content providers 304 and 306 may provide television programs,

advertisements, audio (e.g., radio) programs, still image information (e.g., web pages), etc. in known manners to the digital broadcast station 302. The digital broadcast station 302 transmits one or more signals containing digital program information to a reference site 308 and at least one consumption or home site 310 via communication paths or links 312, 314. The communication paths or links 312 and 314 may be similar to the communication paths or links 14 and 16 (Fig. 1) and, thus, may include any combination of hardwired or wireless links such as, for example, satellite links, wireless land-based links, cable links, etc. The signals conveyed via the links 312 and 314 may contain multi-program data streams, which are commonly employed with existing digital television transmission systems.

[0062] The reference site 308 and the consumption or home site 310 receive and process the digital signals provided by the digital broadcast station 302 using the program identification apparatus and methods described herein. More specifically, the reference site 308 includes a plurality of decoders (e.g., set-top boxes or the like) 316, 318 and 320 that demodulate, demultiplex and decode audio, video and/or other data packets received from the digital broadcast station 302. In an example, each of the decoders 316, 318 and 320 provides data packets associated with a different program that is currently being broadcast to a reference site processor 322. In other words, the digital decoder 316 may provide data packets associated with a program A while the decoders 318 and 320 provide data packets associated with respective programs B and C. In any case, the reference site processor 322 is configured to control and/or has information indicating to which program each of the decoders 316, 318 and 320 is currently tuned.

[0063] The reference site processor 322 includes the apparatus and methods described herein for generating reference signature information for a plurality of

simultaneously broadcast television programs. In particular, the reference site processor 322 may include the signature generator 28, the time stamper 30 and the program identification stamper 32, all of which are shown and described in connection with Fig. 1. In addition, the reference site processor 322 can be implemented to use any of the reference signature generation methods described herein such as, for example, the method described in connection with Figs. 4 and 5. The reference site processor 322 sends the reference signature information it generates to a central processing facility 324, which stores the reference signature information in a database 326 using, for example, the data structure 170 of Fig. 5.

[0064] The consumption site 310 could be, for example, a statistically selected home containing a digital television, a digital radio, a computer, etc. The consumption site 310 includes an output unit 328 (e.g., a video display, speaker, etc.) The consumption site 310 also includes a decoder (e.g., a set-top box) 330, which may be similar or identical to the decoders 316-320. Such decoders are well-known and, thus, are not described in greater detail herein. The decoder 330 provides audio and/or video signals 334 to the output unit 328 that are used to output the program currently selected for consumption. In addition, the decoder 330 also provides signals containing information 336 associated with a currently tuned program to the site unit 332. Such signals may, for example, be audio signals provided using the well-known Sony/Phillips Digital Interface Format (S/PDIF), or using any other desired format that provides data packets associated with digital broadcasts.

[0065] The site unit 332 processes the signals 336 received from the decoder 330 to generate signature information therefrom and conveys the signature information to the central processing facility 324. Thus the site unit 332 includes a signature generator such as, for example, the signature generator 38, and a time stamper such

as, for example, the time stamper 40, both of which may be used as described in connection with Figs. 6 and 7 to generate consumption site signature information.

[0066] The central processing facility 324, among other tasks, is configured to perform the functions of the program consumption analyzer 22 discussed in connection with Figs. 1 and 2. Thus, the central processing facility 324 compares signature information generated at the consumption site 328 to the reference signature information stored in the database 326 to identify the programs consumed at the consumption site 310. The central processing facility 324 may additionally analyze the program consumption information it generates for the consumption site 310 and, possibly, other consumption sites (not shown), to ascertain the consumption behaviors at those consumption sites and/or to develop ratings.

[0067] Although certain methods, apparatus and articles of manufacture have been described herein, the scope of coverage of this patent is not limited thereto. To the contrary, this patent covers all embodiments including apparatus, methods and articles of manufacture fairly falling within the scope of the appended claims either literally or under the doctrine of equivalents.

What is claimed is:

1. A method of identifying a digital data stream having at least one of audio content and video content, comprising:

receiving first signature information associated with a content consumption site, wherein the first signature information is generated using digital information extracted from predetermined portions of the digital data stream;

comparing the first signature information to second signature information associated with a reference site; and

identifying the digital data stream based on the comparison of the first and second signature information.

2. A method as defined in claim 1, wherein each of the predetermined portions of the digital data stream are computed based on another portion of the digital data stream

3. A method as defined in claim 1, wherein the predetermined portions of the digital data stream are associated with non-payload portions of the digital data stream.

4. A method as defined in claim 3, wherein the non-payload portions of the digital data stream include a checksum.

5. A method as defined in claim 4, wherein the checksum is a cyclical redundancy check.

6. A method as defined in claim 1, wherein the predetermined portions of the digital data stream are extracted from at least one of an audio data packet and a video data packet.

7. A method as defined in claim 1, wherein the first signature information includes a plurality of digital values associated with a plurality of data packets extracted from the digital data stream.

8. A method as defined in claim 7, wherein each of the plurality of digital values is generated by combining information extracted from at least two of the plurality of data packets.

9. A method as defined in claim 8, wherein combining the information extracted from the at least two of the plurality of data packets includes concatenating the information extracted from the at least two of the plurality of data packets.

10. A method as defined in claim 1, further comprising receiving a time value associated with the digital information.

11. A method as defined in claim 1, further comprising determining if at least a portion of the first signature information matches at least a portion of the second signature information.

12. A method as defined in claim 11, wherein the at least one portion of the first signature information is a first digital value and the at least one portion of the second signature information is a second digital value.

13. A method as defined in claim 1, wherein comparing the first and second signature information includes comparing the first signature information to digital signatures associated with a plurality of digital data streams, each digital data stream having at least one of digital audio content and digital video content.

14. A method as defined in claim 13, wherein the digital signatures associated with the plurality of digital data streams are generated at the reference site.

15. A method as defined in claim 1, wherein comparing the first and second signature information includes using time values to select a relevant portion of the second signature information.

16. A method as defined in claim 1, wherein the digital data stream is at least one of a digital television program and a digital radio program.

17. A method as defined in claim 1, wherein the digital data stream is in a format suitable for transmission via the Internet.

18. A method as defined in claim 1, wherein the audio content is in a format that is compliant with an audio transmission standard.

19. A method as defined in claim 18, wherein the audio transmission standard is compliant with at least one of AC-3, MPEG, and MPEG advanced audio coding.

20. An apparatus for identifying a digital data stream having at least one of audio content and video content, comprising:

a program consumption analyzer programmed to:

receive first signature information associated with a content consumption site, wherein the first signature information is generated using digital information extracted from predetermined portions of the digital data stream;

compare the first signature information to second signature information associated with a reference site; and

identify the digital data stream based on the comparison of the first and second signature information.

21. An apparatus as defined in claim 20, wherein each of the predetermined portions of the digital data stream are computed based on another portion of the digital data stream

22. An apparatus as defined in claim 20, wherein the digital information is extracted from at least one of an audio data packet and a video data packet.

23. An apparatus as defined in claim 20, wherein the first signature information includes a plurality of digital values associated with a plurality of data packets extracted from the digital data stream.

24. An apparatus as defined in claim 23, wherein each of the plurality of digital values is generated by combining information extracted from at least two of the plurality of data packets.

25. An apparatus as defined in claim 24, wherein combining the information extracted from the at least two of the plurality of data packets includes concatenating the information extracted from the at least two of the plurality of data packets.

26. An apparatus as defined in claim 20, wherein the digital information includes a checksum.

27. An apparatus as defined in claim 26, wherein the checksum is a cyclical redundancy check.

28. An apparatus as defined in claim 20, wherein the first signature information includes a time value associated with the digital information.

29. An apparatus as defined in claim 20, wherein the program consumption analyzer determines if at least a portion of the first signature information matches at least a portion of the second signature information.

30. An apparatus as defined in claim 29, wherein the at least one portion of the first signature information is a first digital value and the at least one portion of the second signature information is a second digital value.

31. An apparatus as defined in claim 20, wherein the program consumption analyzer compares the first and second signature information by comparing the first signature information to digital signatures associated with a plurality of broadcast programs.

32. An apparatus as defined in claim 31, wherein the digital signatures are generated at a reference site.

33. An apparatus as defined in claim 20, wherein the program consumption analyzer compares the first and second signature information using time values to select a relevant portion of the second signature information.

34. An apparatus as defined in claim 20, wherein the digital data stream includes at least one of a digital television program, and a digital radio program.

35. A method as defined in claim 20, wherein the digital data stream is in a format suitable for transmission via the Internet.

36. An apparatus as defined in claim 20, wherein the audio content is in a format that is compliant with an audio transmission standard.

37. An apparatus as defined in claim 36, wherein the audio transmission standard is compliant with at least one of AC-3, MPEG and MPEG advanced audio coding.

38. A machine readable medium having instructions stored thereon that, when executed, cause a machine to:

receive first signature information associated with a content consumption site, wherein the first signature information is generated using digital information extracted from predetermined portions of a digital data stream associated with the content;

compare the first signature information to second signature information associated with a reference site; and

identify the content based on the comparison of the first and second signature information.

39. A machine readable medium as defined in claim 38, wherein the content is audio content.

40. A machine readable medium as defined in claim 38, wherein the content is video content.

41. A machine readable medium as defined in claim 38, wherein each of the predetermined portions of the digital data stream are computed based on another portion of the digital data stream.

42. A machine readable medium as defined in claim 38, wherein the digital information is extracted from at least one of an audio data packet and a video data packet.

43. A machine readable medium as defined in claim 38, wherein the first signature information includes a plurality of digital values associated with a plurality of data packets extracted from the content.

44. A machine readable medium as defined in claim 43, wherein each of the plurality of digital values is generated by combining information extracted from at least two of the plurality of data packets.

45. A machine readable medium as defined in claim 44, wherein combining the information extracted from the at least two of the plurality of data packets includes concatenating the information extracted from the at least two of the plurality of data packets.

46. A machine readable medium as defined in claim 38, wherein the digital information includes a checksum.

47. A machine readable medium as defined in claim 46, wherein the checksum is a cyclical redundancy check.

48. A machine readable medium as defined in claim 38 having instructions stored thereon that, when executed, cause the machine to receive a time value associated with the digital information.

49. A machine readable medium as defined in claim 38 having instructions stored thereon that, when executed, cause the machine to determine if at least a portion of the first signature information matches at least a portion of the second signature information.

50. A machine readable medium as defined in claim 49, wherein the at least one portion of the first signature information is a first digital value and the at least one portion of the second signature information is a second digital value.

51. A machine readable medium as defined in claim 38 having instructions stored thereon that, when executed, cause the machine to compare the first and second signature information by comparing the first signature information to digital signatures associated with a plurality of digital data streams.

52. A machine readable medium as defined in claim 51, wherein the digital signatures are generated at the reference site.

53. A machine readable medium as defined in claim 38 having instructions stored thereon that, when executed, cause the machine to compare the first and second signature information using time values to select a relevant portion of the second signature information.

54. A machine readable medium as defined in claim 38, wherein the content includes at least one of a digital television program, and a digital radio program.

55. A machine readable medium as defined in claim 38, wherein the content includes content that is in a format that is suitable for transmission via the Internet.

56. A machine readable medium as defined in claim 38, wherein the content is in a format that is compliant with an audio transmission standard.

57. A machine readable medium as defined in claim 56, wherein the audio transmission standard is compliant with at least one of AC-3, MPEG and MPEG advanced audio coding.

58. A method of generating digital data stream signatures, comprising:

acquiring a plurality of data packets associated with a digital data stream;

extracting digital values from at least some of the plurality of data packets; and

generating a plurality of signatures based on the digital values.

59. A method as defined in claim 58, further comprising:

associating at least one of an identifier indicative of a content contained in the digital data stream and a time value with the plurality of signatures.

60. A method as defined in claim 58, wherein each of the digital values is computationally related to a corresponding one of the plurality of data packets.

61. A method as defined in claim 58, wherein the digital data stream includes audio content and wherein each of the plurality of data packets is associated with the audio content.

62. A method as defined in claim 61, wherein each of the plurality of data packets is formatted according to an audio transmission standard.

63. A method as defined in claim 62, wherein the audio transmission standard is compliant with at least one of AC-3, MPEG and MPEG advanced audio coding.

64. A method as defined in claim 58, wherein the digital data stream includes content comprising at least one of a digital television program and a digital radio program.

65. A method as defined in claim 58, wherein the digital data stream is formatted for transmission via the Internet.

66. A method as defined in claim 58, wherein extracting the digital values from the at least some of the plurality of data packets includes extracting checksums from the at least some of the plurality of data packets.

67. A method as defined in claim 58, wherein generating the plurality of signatures based on the digital values includes combining at least portions of the digital values to generate the signatures.

68. A method as defined in claim 67, wherein combining the at least portions of the digital values to generate the signatures includes concatenating the at least portions of the digital values.

69. A method as defined in claim 58, wherein the digital values are checksums.

70. An apparatus for generating signatures, comprising:

a signal processor that acquires at least one data packet associated with a content comprising at least one of digital audio content and digital video content; and

a signature generator that extracts values from the at least one data packet and generates a signature based on the extracted values.

71. An apparatus as defined in claim 70, wherein each of the extracted values is computationally related to a corresponding one of the at least one data packet.

72. An apparatus as defined in claim 70 including a time stamper that associates time values with the extracted values.

73. An apparatus as defined in claim 70 including a program stamper that associates an identifier indicative of the content with the signature.

74. An apparatus as defined in claim 70, wherein the at least one data packet is associated with audio information contained in the content.

75. An apparatus as defined in claim 70, wherein the at least one data packet is formatted to comply with an audio transmission standard.

76. An apparatus as defined in claim 75, wherein the audio transmission standard is compliant with at least one of AC-3, MPEG and MPEG advanced audio coding.

77. An apparatus as defined in claim 70, wherein the content is a digital television program.

78. An apparatus as defined in claim 70, wherein the signature generator is configured to extract the values from the at least one data packet by extracting a checksum from the data packet.

79. An apparatus as defined in claim 70, wherein the signature generator is configured to generate the signature based on the extracted values by combining at least portions of the extracted values to generate the signature.

80. An apparatus as defined in claim 79, wherein the signature generator is configured to combine the at least portions of the extracted values to generate the signatures by concatenating the at least portions of the extracted values.

81. An apparatus as defined in claim 80, wherein the extracted values are checksums.

82. A method of identifying a digital television program, comprising:

receiving consumption site signature information associated with the digital television program from a consumption site;

comparing the consumption site signature information to reference site signature information associated with the digital television program, wherein the consumption site signature information and the reference site signature information are generated using predetermined data packets; and

identifying the digital television program based on the comparison of the consumption site signature information and the reference site signature information.

83. A method as defined in claim 1, wherein the predetermined data packets are computationally related to other data packets.

84. A method as defined in claim 1, wherein receiving the consumption site signature information includes receiving a plurality of digital values and time values associated with the digital values.

85. A method as defined in claim 84, wherein the digital values are generated using checksums extracted from a digital television signal.

86. A method as defined in claim 1, including determining if at least some of the consumption site signature information matches at least some of the reference signature site information.

87. A method as defined in claim 1, wherein identifying the digital television program based on the comparison of the consumption site signature information and the reference site signature information includes using a program identifier associated with the reference site signature information to identify the digital television program.

88. An apparatus for identifying a digital television program, comprising:

a processing site; and

a database coupled to the processing site, wherein the processing site is programmed to:

receive consumption signature information associated with the digital television program from a consumption site, wherein the consumption signature information is generated based on one or more predetermined data packets;

compare the consumption signature information to reference signature information associated with the digital television program; and

identify the digital television program based on the comparison of the consumption signature information and the reference signature information.

89. An apparatus as defined in claim 88, wherein at least portions of the predetermined data packets are computationally related to at least portions of other data packets.

90. An apparatus as defined in claim 88, wherein the processing site is programmed to receive the consumption signature information including a plurality of digital values and time values associated with the digital values.

91. An apparatus as defined in claim 90, wherein the digital values are generated using checksums extracted from a digital television signal.

92. An apparatus as defined in claim 88, wherein the processing site is programmed to determine if at least some of the consumption signature information matches at least some of the reference signature information.

93. An apparatus as defined in claim 88, wherein the processing site uses a program identifier associated with the reference signature information to identify the digital television program.

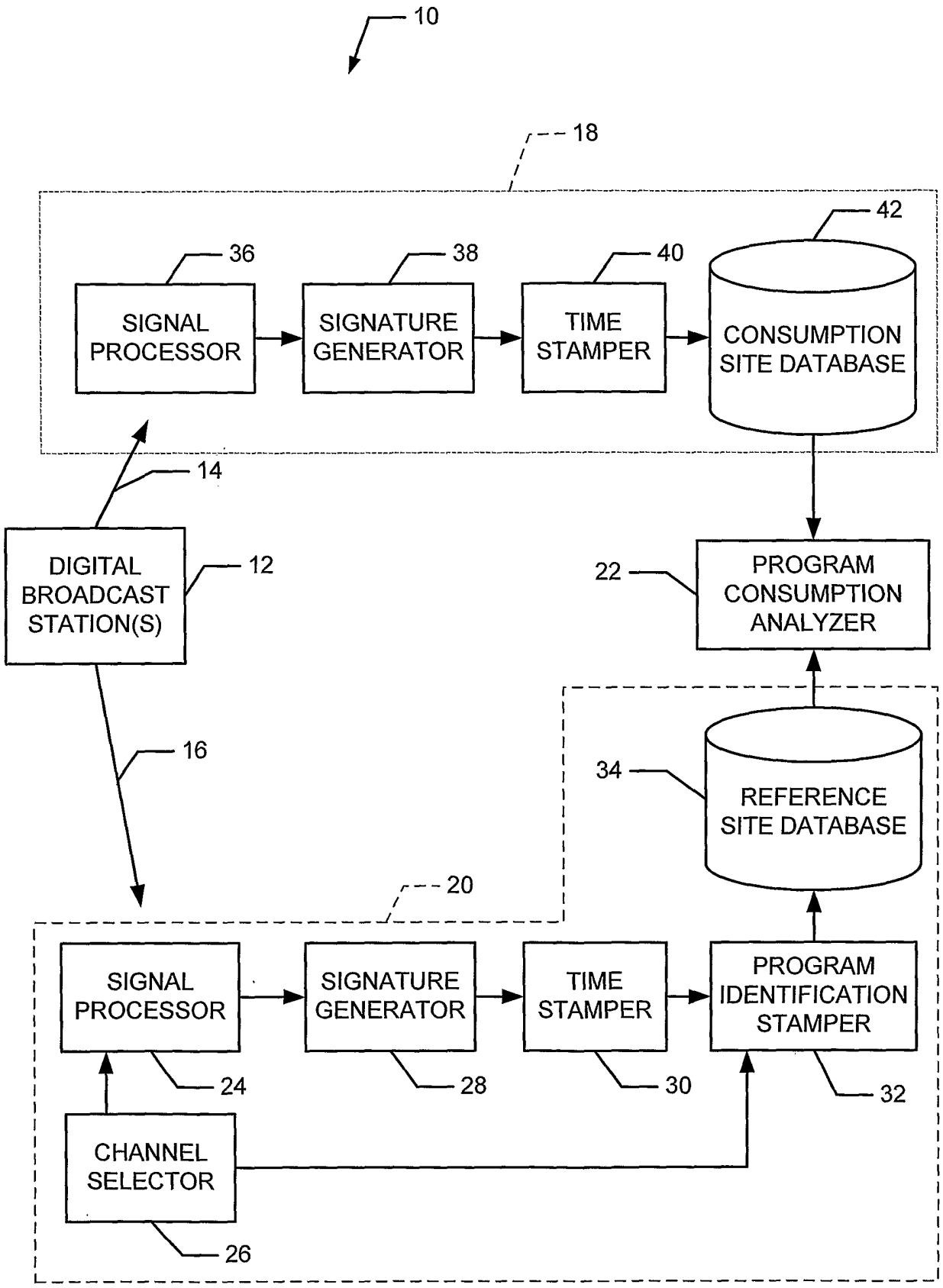

10
18
42
36
38
40
SIGNAL PROCESSOR
SIGNATURE GENERATOR
TIME STAMPER
CONSUMPTION SITE DATABASE
14
DIGITAL BROADCAST STATION(S)
12
22
PROGRAM CONSUMPTION ANALYZER
16
34
REFERENCE SITE DATABASE
20
SIGNAL PROCESSOR
SIGNATURE GENERATOR
TIME STAMPER
PROGRAM IDENTIFICATION STAMPER
24
28
30
32
CHANNEL SELECTOR
26

FIG. 1

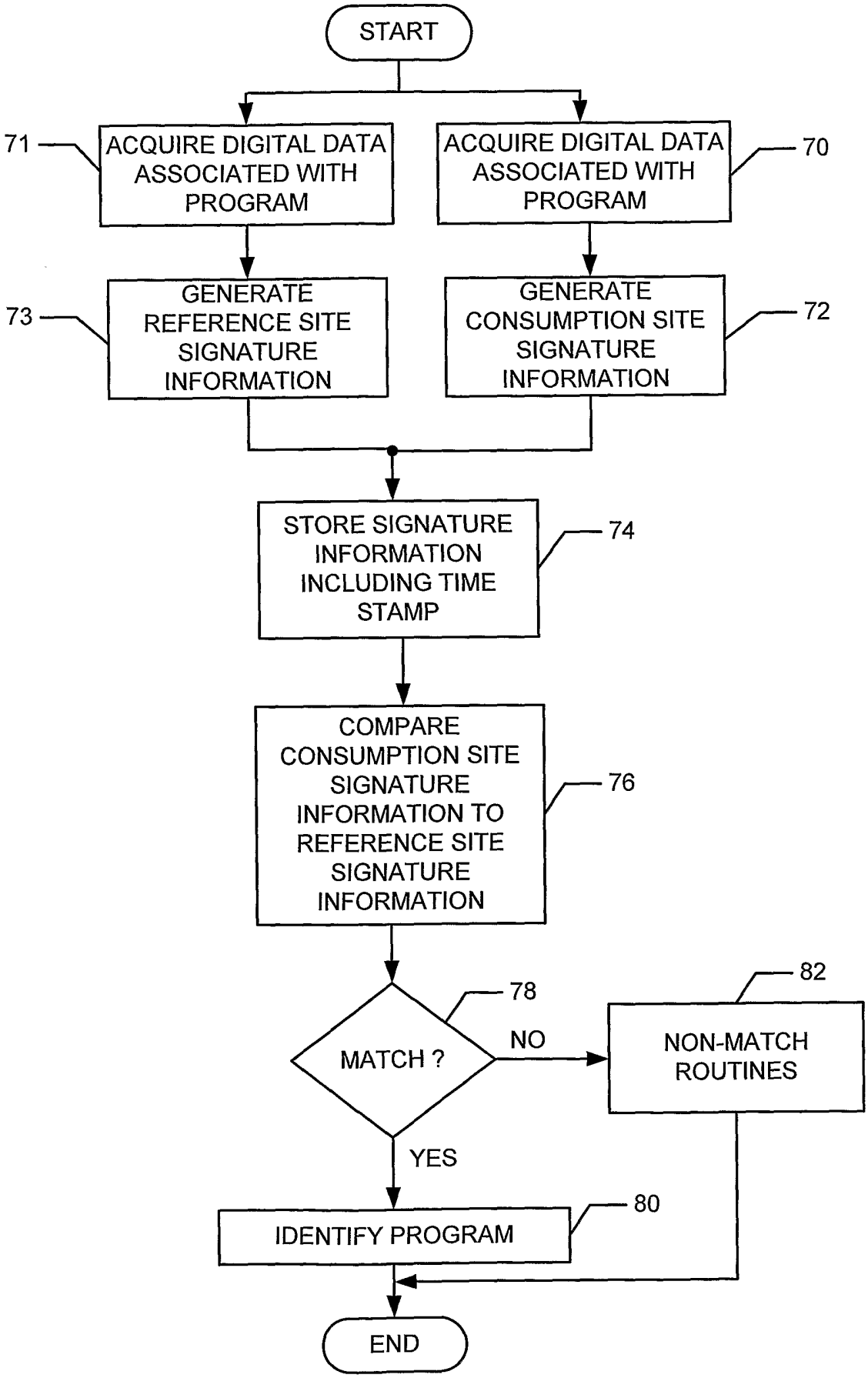

START
71
ACQUIRE DIGITAL DATA ASSOCIATED WITH PROGRAM
ACQUIRE DIGITAL DATA ASSOCIATED WITH PROGRAM
70
73
GENERATE REFERENCE SITE SIGNATURE INFORMATION
GENERATE CONSUMPTION SITE SIGNATURE INFORMATION
72
STORE SIGNATURE INFORMATION INCLUDING TIME STAMP
74
COMPARE CONSUMPTION SITE SIGNATURE INFORMATION TO REFERENCE SITE SIGNATURE INFORMATION
76
78
MATCH ?
NO
82
NON-MATCH ROUTINES
YES
80
IDENTIFY PROGRAM
END

FIG. 2

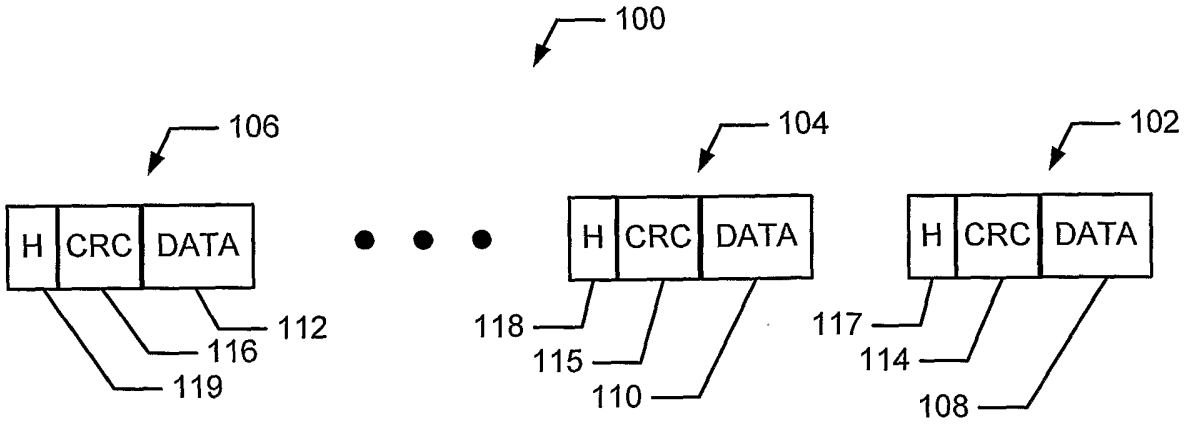


FIG. 3

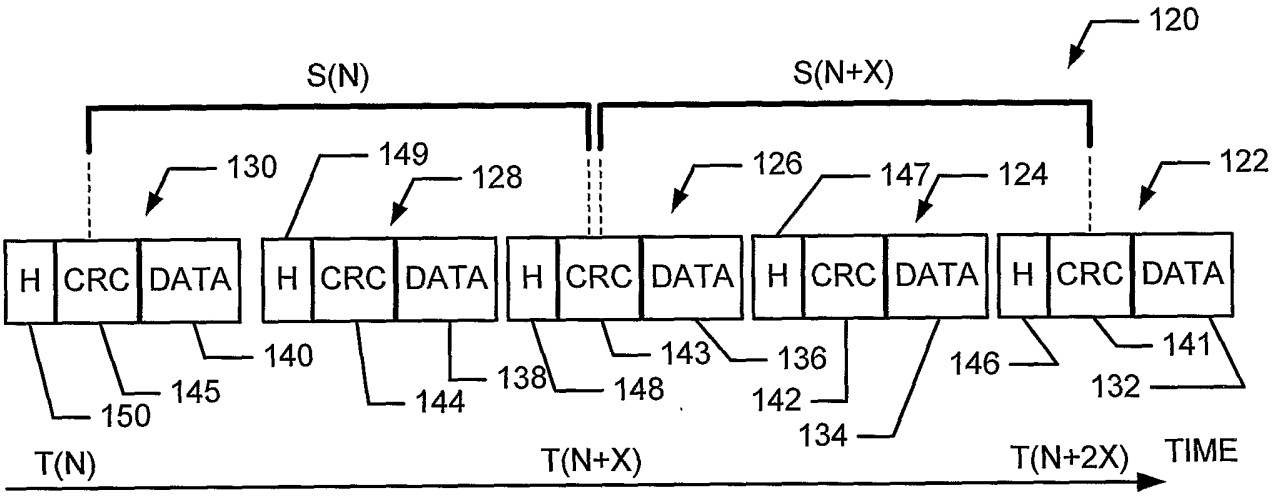
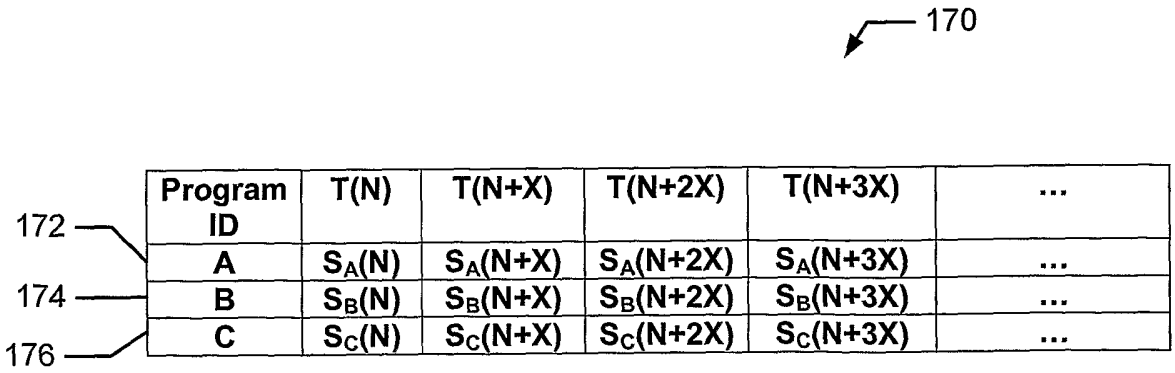


FIG. 4



Program ID	T(N)	T(N+X)	T(N+2X)	T(N+3X)	...
A	$S_A(N)$	$S_A(N+X)$	$S_A(N+2X)$	$S_A(N+3X)$	...
B	$S_B(N)$	$S_B(N+X)$	$S_B(N+2X)$	$S_B(N+3X)$	...
C	$S_C(N)$	$S_C(N+X)$	$S_C(N+2X)$	$S_C(N+3X)$	...

FIG. 5

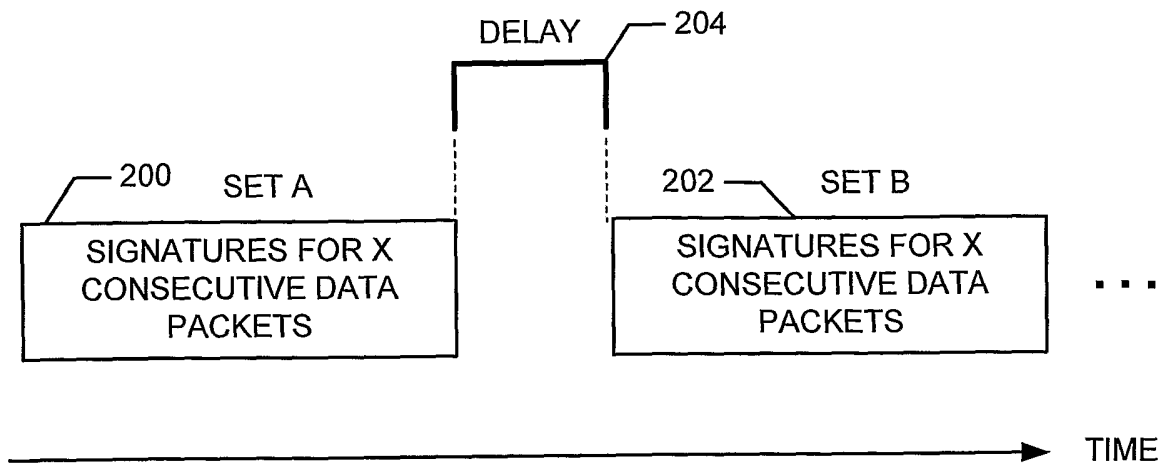

DELAY
204
200
SET A
202
SET B
SIGNATURES FOR X CONSECUTIVE DATA PACKETS
SIGNATURES FOR X CONSECUTIVE DATA PACKETS
. . .
TIME

FIG. 6

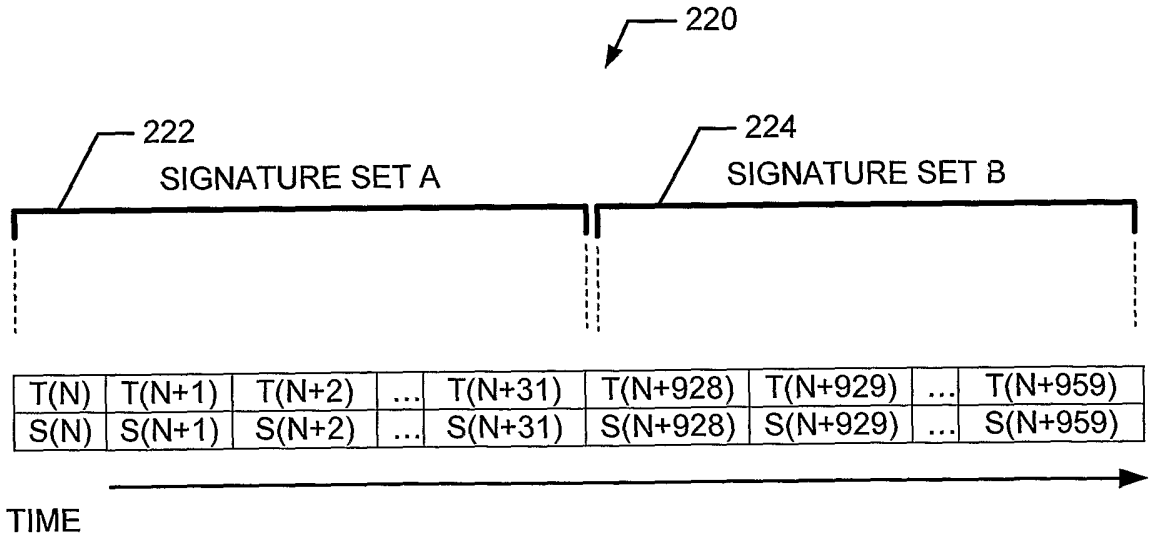

220
222
SIGNATURE SET A
224
SIGNATURE SET B
T(N) T(N+1) T(N+2) ... T(N+31) T(N+928) T(N+929) ... T(N+959)
S(N) S(N+1) S(N+2) ... S(N+31) S(N+928) S(N+929) ... S(N+959)
TIME

FIG. 7

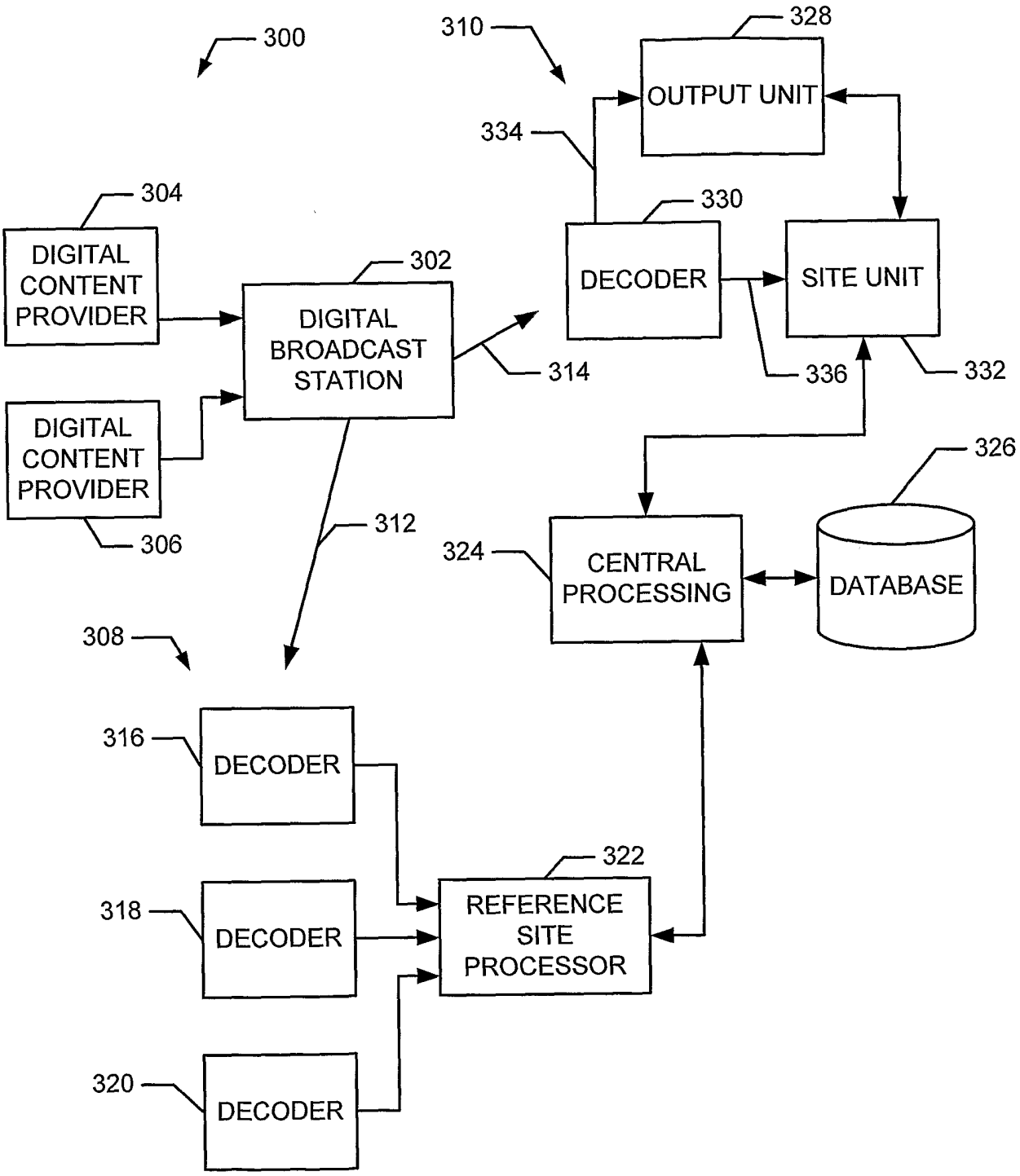

300
310
328
OUTPUT UNIT
334
304
330
DIGITAL CONTENT PROVIDER
302
DIGITAL BROADCAST STATION
DECODER
SITE UNIT
314
336
332
DIGITAL CONTENT PROVIDER
326
306
312
324
CENTRAL PROCESSING
DATABASE
308
316
DECODER
322
318
DECODER
REFERENCE SITE PROCESSOR
320
DECODER

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/22562

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : H04N 9/00, 7/16, 7/00
US CL : 725/9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21; 348/465

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 725/9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21; 348/465

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2002/0010919 A1 (Lu et al.) 24 January 2002 (24.01.2002), [0031-0052]	1-3,6-11,14,16-19,20-25,28,29,32,34-36-38,40-45,48,49,52,54-56-58-64,65,67,68,70-77,79,80 ---------- 4,5,12,13,15,17,19,26,27,30,31,33,35,37,39,46,47,50,51,53,55,57,63,65,66,69,76,78,81
X --- Y	US 2003/0005430 A1 (Kolessar) 02 January 2003 (02.01.2003), [0030-0056]	82-84,86,87,88,89,90,92,93 ---------- 4,5,26,27,46,47,78,81,85,91

☒ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search 31 December 2003 (31.12.2003)	Date of mailing of the international search report 22 JAN 2004
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US Commissioner for Patents P.O. Box 1450 Alexandria, Virginia 22313-1450 Facsimile No. (703) 305-3230	Authorized officer Johnny Ma Telephone No. (703) 308-HELP

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

C. (Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 4,739,398 (Thomas et al.) 19 April 1988 (19.04.1988), columns 3-10	1,2,6-9,11,20-25,29,38,39,40,41-45,49 ---------- 39
Y	US 5,621,454 (Ellis et al.) 15 April 1997 (15.04.1997), column 10, lines 31-50	2,21,41,60,71,83,89
Y	US 5,481,294 (Thomas et al.) 02 January 1996 (02.01.1996), column 15, lines 1-55	12,13,15,30,31,33,50,51,53,72
Y	US 5,689,509 (Gaytan et al.) 18 November 1997 (18.11.1997), column 4, lines 27-41	
Y	US 5,404,160 (Schober et al.) 04 April 1995 (04.04.1995), column 3, lines 55-63; column 4, lines 37-44.	87,93
A	US 5,485,518 (Hunter et al.) 16 January 1996 (16.01.1996)	
A	US 6,577,346 B1 (Perlman) 10 June 2003 (10.06.2003)	
A	US 4,677,466 (Lert, Jr. et al.) 30 June 1987 (30.06.1987)	
A	US 6,112,053 (Dunki-Jacobs et al.) 29 August 2000 (29.08.2000)	
A	US 2002/0186768 (Dimitrova et al.) 12 December 2002 (12.12.2002)	
A	US 5,974,299 (Massetti) 26 October 1999 (26.12.1999)	
A	US 6,469,749 B1 (Dimitrova et al.) 22 October 2002 (22.10.2002)	
A	US 5,019,899 (Boles et al.) 28 May 1991 (28.05.1991)	
A	US 2003/0023967 A1 (Kim) 30 January 2003 (30.01.2003)	